Claims

What is claimed is:

1. A method of person tagging in an image processing system, the method comprising the steps of:

processing a sequence of images to generate a statistical model for each person to be tagged, the statistical model incorporating at least one appearance feature and at least one geometric feature of the tagged person;

applying the model to at least one subsequent image in order to perform at least one of a detection operation, a location operation and a tracking operation for the tagged person; and

controlling an action of the image processing system based on a result of the at least one operation.

2. The method of claim 1 wherein the sequence of images comprises a video segment.

3. The method of claim 1 wherein the processing step further includes processing the sequence of images to generate a plurality of statistical models, each of the models corresponding to a particular tagged person.

4. The method of claim 1 wherein the appearance feature comprises at least one of a color feature and a texture feature.

5. The method of claim 1 wherein the geometric feature comprises at least one of a region shape and a region position for a given one of a plurality of regions associated with the statistical model.

6. The method of claim 1 wherein the statistical model is generated at least in part by segmenting a given image into a number *N* of different regions of similar appearance.

7. The method of claim 1 wherein the statistical model generated for a given person Ω comprises a likelihood probability function $P(\mathbf{I} \mid \Omega$ which indicates the likelihood that the person Ω is present in a given image **I**.

8. The method of claim 7 wherein the likelihood probability function $P(\mathbf{I} \mid \Omega$ for person Ω is computed as

$$P(\mathbf{I}|\Omega) = \sum_{r=1,2,\ldots N} P(R_r|\Omega)P(r|\Omega ,$$

where R_r is a function of the at least one appearance feature and the at least one geometric feature, and r is an index identifying one of N regions of similar appearance within the image **I**.

9. The method of claim 1 wherein the statistical model generated for a given person Ω comprises a likelihood probability function $P(\mathbf{I} \mid \mathbf{T}, \xi, \Omega$, where **T** is a linear transformation used to capture global motion of the person in an image **I**, and ξ is a discrete variable used to capture local motion of the person at a given point in time.

10. The method of claim 9 wherein a location of the person is determined using the linear transformation **T**.

11. The method of claim 9 wherein a pose of the person is determined using the discrete variable ξ.

12. The method of claim 9 wherein the linear transformation **T** is used to obtain a sub-window of the image **I** that is invariant to rotation and scale.

13. The method of claim 9 wherein the linear transformation **T** is

implemented using a bilinear interpolation technique with a reference point $\mathbf{x}_c$ in the image $\mathbf{I}$, a rotation angle θ, and a scaling factor s.

14. The method of claim 9 wherein the local motion is modeled using a discrete set of states $\{\xi_1, \xi_2, \ldots \xi_M$ of the variable ξ to capture M different poses of the person Ω.

15. The method of claim 1 wherein the statistical model generated for a given person Ω and image $\mathbf{I}$ comprises a likelihood probability function

$$P(\mathbf{I} \mid \mathbf{T}, \xi, \Omega) = \sum_{pix \in \mathbf{I}} P(pix \mid \mathbf{T}, \xi, \Omega ,$$

where r is an index to regions of similar appearance and N is a total number of such regions, $r = 1, 2, \ldots N$, and

$$P(pix \mid \mathbf{T}, \xi, \Omega) = \max_{r=1,\ldots,N} \left[P(pix \mid r, \mathbf{T}, \xi, \Omega) \, P(r \mid \xi, \Omega) \right],$$

where $P(pix \mid r, \mathbf{T}, \xi, \Omega$ is the probability of observing pixel pix assuming that it belongs to an r-th region of the model on a pose ξ, and $P(r \mid \xi, \Omega$ is the prior probability of the region at that pose.

16. The method of claim 15 wherein the regions of similar appearance include a dummy region having a constant probability as follows:

$$P(pix \mid r_{occlusion}, \mathbf{T}, \xi, \Omega) \, P(r_{occlusion} \mid \xi, \Omega) = P_{occlusion} .$$

17. The method of claim 15 wherein each of at least a subset of

the pixels of the image **I** is characterized by a two-dimensional position vector **x** and by an appearance feature vector **f** such that:

$$P(pix \mid r, \mathbf{T}, \xi, \Omega) = P(\mathbf{x} \mid r, \mathbf{T}, \xi, \Omega)\, P(\mathbf{f} \mid r, \mathbf{T}, \xi, \Omega ,$$

where $P(\mathbf{x} \mid r, \mathbf{T}, \xi, \Omega$ and $P(\mathbf{f} \mid r, \mathbf{T}, \xi, \Omega$ are approximated as Gaussian distributions over corresponding feature spaces.

18. The method of claim 1 wherein the controlling step comprises generating an output of the image processing system based on the result of the at least one operation.

19. The method of claim 1 wherein the controlling step comprises altering an operating parameter of the image processing system based on the result of the at least one operation.

20. An apparatus for use in providing person tagging in an image processing system, the apparatus comprising:

a processor operative to process a sequence of images to generate a statistical model for each person to be tagged, the statistical model incorporating at least one appearance feature and at least one geometric feature of the tagged person, the processor being further operative to apply the model to at least one subsequent image in order to perform at least one of a detection operation, a location operation and a tracking operation for the tagged person, and further wherein an action of the image processing system is controlled based on a result of the at least one operation.

21. An article of manufacture comprising a storage medium for storing one or more programs for use in providing person tagging in an image processing system, wherein the one or more programs when executed by a processor implement the steps of:

processing a sequence of images to generate a statistical

model for each person to be tagged, the statistical model incorporating at least one appearance feature and at least one geometric feature of the tagged person; and

applying the model to at least one subsequent image in order to perform at least one of a detection operation, a location operation and a tracking operation for the tagged person;

wherein an action of the image processing system is controlled based on a result of the at least one operation.